Table of Contents

[Executive summary 3](#_Toc154015396)

[Introduction 3](#_Toc154015397)

[Key Findings 3](#_Toc154015398)

[Recommendations 3](#_Toc154015399)

[Personal Conclusions 3](#_Toc154015400)

[Introduction 3](#_Toc154015401)

[Dataset Description 3](#_Toc154015402)

[BI Requirements 3](#_Toc154015403)

[Why Did I Choose This Dataset? 3](#_Toc154015404)

[Does My Dataset Address the Big Data Problem? 3](#_Toc154015405)

[Which specific features are you going to focus on? 3](#_Toc154015406)

[Will this dataset help you in developing specific business skills? 3](#_Toc154015407)

[What questions do you seek to answer with your BI project? 3](#_Toc154015408)

[Data Preprocessing 3](#_Toc154015409)

[Download the Data Sets 3](#_Toc154015410)

[Combine Multiple Data Sets 3](#_Toc154015411)

[Create Single Table 3](#_Toc154015412)

[Removing Duplicates 3](#_Toc154015413)

[Removing Empty Rows 3](#_Toc154015414)

[Removing Blank Rows and Null Data 3](#_Toc154015415)

[Changing Data Types 3](#_Toc154015416)

[Removing Columns 3](#_Toc154015417)

[Renaming Columns 3](#_Toc154015418)

[Column Manipulation 3](#_Toc154015419)

[Calculated Columns 3](#_Toc154015420)

[Calculated Measure 3](#_Toc154015421)

[Data Model 4](#_Toc154015422)

[Normalization 4](#_Toc154015423)

[Street Crime 4](#_Toc154015424)

[Location 4](#_Toc154015425)

[Crime Type 4](#_Toc154015426)

[Report 4](#_Toc154015427)

[Create Relationship 4](#_Toc154015428)

[Data Visualization 4](#_Toc154015429)

[Dashboards 4](#_Toc154015430)

[A 4](#_Toc154015431)

[B 4](#_Toc154015432)

[Appearance and navigation 4](#_Toc154015433)

[Custom Theme Design 4](#_Toc154015434)

[Slicer Panel 4](#_Toc154015435)

[Navigation 4](#_Toc154015436)

[References 4](#_Toc154015437)

# Executive summary

## Introduction

## Key Findings

## Recommendations

## Personal Conclusions

# Introduction

To understand the crime dynamics within the northeast of England, particularly in areas like Newcastle, Middlesbrough, Redcar and others, I have embarked on a Business Intelligence (BI) project using data obtained from the [data.Police.uk](https://data.police.uk/data/) website. This project aims to delve into crime patterns in these regions, with a special focus on street crimes.

## Dataset Description

The dataset spans from January to October 2023 and covers two police jurisdictions: Cleveland and Northumbria Police. It is comprehensive, including columns such as Crime ID, Month, Reporting Agency, Geographic Coordinates (Longitude, Latitude), Location, Lower Layer Super Output Area (LSOA) code and name, Crime Type, Last Outcome Category, and Context.

## BI Requirements

### Why Did I Choose This Dataset?

I chose this dataset primarily because of my personal interest in the safety of the area where I live. The Northeast, particularly cities like Middlesbrough and Redcar, has a reputation for street crimes. This dataset provides an opportunity to analyze crime trends and patterns in these areas.

### Does My Dataset Address the Big Data Problem?

The dataset addresses the big data problem by providing a large volume of varied data over a significant time span. It allows for a detailed analysis of crime types, locations, and their changes over time.

### Which specific features are you going to focus on?

I will focus on crime types, specifically street crimes, and their geographical distribution within the cities. The dataset's longitudinal aspect also allows for the examination of temporal trends in crime rates.

### Will this dataset help you in developing specific business skills?

Yes, analyzing this dataset will enhance my skills in data analytics, pattern recognition, and geographical information systems (GIS). It will also improve my ability to draw actionable insights from complex data sets.

### What KPI questions do you seek to answer with your BI project?

# Data Preprocessing

## Download the Data Sets

The initial step involved downloading separate files for each month from the Police.co.uk website, for both Cleveland and Northumbria Police departments.

## Combine Multiple Data Sets

To streamline the analysis, these monthly files were combined into a single dataset. This was accomplished using Power BI’s Combine Load to select a folder of files and combine them upon loading, other option I have gone through merge and append functionalities, but I choose to select the combine and select folder option, one more thing we can achieve it by placing the upcoming data sets file that are not published yet, once it will published we can place the file in data sets folder and we can Scheduled Refresh In Power BI Service, we can set up a schedule to refresh the dataset at specific intervals. This is useful for keeping our data up to date without manual intervention. The frequency of refresh can vary depending on your Power BI service plan.

## Create Single Table

After combining the files, a single table was created, named "Street Crime," to facilitate a unified analysis.

## Removing Duplicates

Duplicates in the dataset were identified and removed. This step is crucial to ensure the accuracy and reliability of the analysis.

## Removing Empty Rows

Empty rows, which can skew results and analyses, were identified and eliminated from the dataset.

## Removing Blank Rows and Null Data

Further data cleaning involved the removal of blank rows and null data across all columns, enhancing the dataset's integrity.

## Changing Data Types

Each column’s data type was examined and adjusted as necessary to align with the nature of the data it contained.

## Removing Columns

Columns that were deemed irrelevant, such as 'Source.Name' and 'Context', were removed. This step streamlined the dataset for more focused analysis.

## Renaming Columns

The 'Month' column was renamed to 'Date' for better clarity and to reflect the nature of the data more accurately.

## 

## Column Manipulation

### Calculated Columns

### Calculated Measure

Several calculated measures were introduced for enhanced analysis:

For the 'CrimeType' table: 'Total Crime by Type' and 'Unique Crime Types'.

For the 'Location' table: 'Highest Crime Type and Count', 'Least Crime Type and Count', 'Total Crime by Location', and 'Unique Locations'. For the 'Report' table: 'Total Report'. For the 'Street Crime' table: 'Count by Crime', 'Crime by Type', 'Percentage', 'Total Crime', and 'Total Outcomes'.

These calculated columns and measures are instrumental in providing a nuanced understanding of the data, allowing for a detailed analysis of street crime trends and patterns in the covered areas.

# Data Model

## The process of normalizing the data involved breaking down the single 'Street Crime' table into three separate tables: 'Location', 'Crime Type', and 'Report'. This approach helps in organizing the data into a more manageable and efficient structure.

## Normalization

Normalization was achieved by following these steps:

### **Street Crime**: Retained essential columns like 'Crime ID', 'Date', 'Falls within', and 'Last outcome category', making it the central fact table that holds the key data points for each crime incident.

### **Location**: This dimension table includes 'Crime ID' (as a foreign key), 'Latitude', 'Longitude', 'Location', 'LSOA code', and 'LSOA name'. These columns contain geographical data that describe where each crime occurred.

### **Crime Type:** Contains 'Crime ID' (as a foreign key) and 'Crime type'. This table categorizes each crime incident, allowing for analysis based on the type of crime.

### **Report:** Includes 'Crime ID' and 'Reported by', linking each crime to the entity that reported it.

By duplicating the original 'Street Crime' table and retaining only the necessary columns for each of the new tables, a clean and normalized set of tables was created.

## 

## Create Relationship

Power BI's auto-detection feature recognized the relationships between these tables based on the 'Crime ID' column, which acts as a primary key in the 'Street Crime' table and a foreign key in the other tables. The relationships are likely one-to-many, with 'Street Crime' being the one side, indicating that each crime ID is unique to the 'Street Crime' table and can be related to multiple entries in the other tables if there are multiple reports, locations, or types associated with a single crime.

The data model diagram showing how the data is connected. Having such a normalized data model is beneficial for several reasons:

A screenshot of a computer

Description automatically generated

Figure 2: Data model diagram after normalization

It reduces redundancy and inconsistency, ensuring that each data element is stored only once.

It simplifies the maintenance of the data because updates, deletions, and insertions are made in just one place.

It enhances data integrity and accuracy, which are crucial for reliable analysis and reporting.

With the relationships established, we can now perform a wide range of analyses using Power BI's powerful data visualization tools. we can explore crime trends over time, analyze crime by location, and compare the frequency of different crime types, among other insights. This normalized data model is a strong foundation for a comprehensive BI solution.

# Data Visualization

## Dashboards

### Crime Overview Dashboard

A screenshot of a computer

Description automatically generated

### Crime Sporting Dashboard

A screenshot of a computer

Description automatically generated

### Area wise Crime Dashboard

A screenshot of a computer

Description automatically generated

### Crime Report Changes Dashboard

A screenshot of a computer

Description automatically generated

## Appearance and navigation

### Custom Theme Design

### Slicer Panel

### Navigation

# References